

Factual Accuracy and Trust in Information: The Role of Expertise

Teun Lucassen and Jan Maarten Schraagen

Department of Cognitive Psychology and Ergonomics, University of Twente, P.O. Box 215, 7500 AE Enschede, The Netherlands. E-mail: {t.lucassen, j.m.c.schraagen}@gw.utwente.nl

In the past few decades, the task of judging the credibility of information has shifted from trained professionals (e.g., editors) to end users of information (e.g., casual Internet users). Lacking training in this task, it is highly relevant to research the behavior of these end users. In this article, we propose a new model of trust in information, in which trust judgments are dependent on three user characteristics: source experience, domain expertise, and information skills. Applying any of these three characteristics leads to different features of the information being used in trust judgments; namely source, semantic, and surface features (hence, the name 3S-model). An online experiment was performed to validate the 3S-model. In this experiment, Wikipedia articles of varying accuracy (semantic feature) were presented to Internet users. Trust judgments of domain experts on these articles were largely influenced by accuracy whereas trust judgments of novices remained mostly unchanged. Moreover, despite the influence of accuracy, the percentage of trusting participants, both experts and novices, was high in all conditions. Along with the rationales provided for such trust judgments, the outcome of the experiment largely supports the 3S-model, which can serve as a framework for future research on trust in information.

Introduction

Since the 1980s, there has been a shift in responsibility for the verification of information credibility. Earlier, this task was mostly performed by professionals. Newspaper editors, for instance, used to decide which pieces of information were suitable for release to the general public. Credibility was one of the decisive factors for this decision, along with, for example, relevance to the public and readability. Nowadays, the task of distinguishing credible information from less credible information often lies with the end user of the information (Flanagin & Metzger, 2007). The introduction of

the World Wide Web (and especially Web 2.0) has resulted in a much larger range of information suppliers than before, for which expert evaluations of credibility are often not available. Online information is not less credible, per se, but users should be aware of the possibility of encountering low-quality information. A good example is Wikipedia: Research has shown that its information quality is overall very high (e.g., Giles, 2005; Rajagopalan et al., 2010), but the open-editing model combined with the tremendous number of articles (>3.4 million¹) requires users to always be aware of the risk of low-quality information (Denning, Horning, Parnas, & Weinstein, 2005).

A highly relevant topic for research is how lay people cope with the varying credibility of information. While they need to make assessments of credibility, they typically are not trained for this task as are professionals. It is suggested in the existing literature that individual differences among users influence trust assessment behavior. In this study, we attempt to explain these differences in terms of user characteristics, particularly focusing on trust in information of Internet users with varying levels of expertise on the topic at hand. This relationship between domain expertise of the user and trust judgments is especially new in the field of information credibility and trust research.

In this article, we first discuss the concept of trust, of which no consensus has been reached by researchers in the various relevant fields. Second, we propose a new model of trust in information. We use this model to predict that various characteristics of a user lead him or her to employ different features of the information to judge its credibility. We then continue to discuss in detail three types of relevant user characteristics: domain expertise, information skills, and source experience. Our hypotheses aim at validating the proposed model. After this, our method using online questionnaires featuring Wikipedia articles with manipulated accuracy is introduced to test the hypotheses. Finally, the results are presented and discussed.

Received September 7, 2010; revised March 11, 2011; accepted March 11, 2011

© 2011 ASIS&T • Published online 19 April 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.21545

¹<http://en.wikipedia.org/wiki/Special:Statistics>

Trust

The concept of trust has been studied in various ways in the literature. Kelton, Fleischmann, and Wallace (2008) distinguished four levels of trust: *individual* (an aspect of personality), *interpersonal* (one actor trusting another), *relational* (an emergent property of a mutual relationship), and *societal* (a characteristic of a whole society). The most common approach of studying trust is at the interpersonal level, concerning a one-way tie between a trustor (someone who trusts) and a trustee (someone who is trusted). An often-used definition of trust at this level was given by Mayer, Davis, and Schoorman (1995):

The willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party. (p. 712)

Trustors assess trustees to determine the degree to which the trustees can be trusted. These assessments often include estimating various characteristics of the trustee, deemed relevant to trustworthiness by the trustor (trust antecedents). These typically include factors such as perceived competence, intentions, and openness.

According to Kelton et al. (2008), interpersonal trust also is the appropriate level to apply to the study of trust in information because information is produced by an author (trustee) and communicated over a certain channel to a receiver (trustor). Assessing trust in the information thus can be seen as assessing trust in the author. However, next to the assessment of characteristics of the author, assessing trust in information also may include characteristics (features) of the information itself. This approach seems especially useful when the author of the information is unknown or when a piece of information has multiple authors. An example of such a case is Wikipedia, where multiple, often anonymous authors contribute to one article. In such situations, the assessment of characteristics of the author(s) may become overly complex or even impossible. Alexander and Tate (1999) identified five criteria that always should be considered when assessing trust in information: accuracy, authority, objectivity, currency, and coverage. Cues of at least four of these criteria (all but authority) also may be found in the information itself, without knowing the identity of the author.

A term often used interchangeably with (information) trust is credibility; however, there is a slight difference. Fogg and Tseng (1999) summarized this difference as credibility meaning believability and as trust meaning dependability. Credibility can be described as perceived information quality, or the assessment of information quality by a user. Credibility is mostly seen as consisting of two key elements: trustworthiness (well-intentioned) and expertise (knowledgeable). Trust, however, also introduces the notion of willingness to depend on the credibility of information. This dependency involves a certain risk that someone takes by using the information (Kelton et al., 2008).

In the remainder of this article, we refer to “trust” as a property of the information user. Credibility is used as the aspect of information that is being considered when judging trust.

A model of online trust proposed by Corritore, Kracher, and Wiedenbeck (2003) has shed more light on the relationship between trust and credibility. Two factors influencing trust were identified in their model: external factors and individual perception. External factors can influence the perception of trust, which in turn is composed of three factors: credibility, ease of use, and risk.

Kelton et al. (2008) proposed an integrated model of trust in information. According to this model, trust also may stem from other factors than the assessment of trustworthiness, such as the disposition to the information, relevance of the information, and recommendations. Personal factors, such as confidence and willingness to trust, also may contribute. This suggests that users with varying (personal) characteristics may judge the same information very differently.

The unifying framework of credibility assessment, as proposed by Hilligoss and Rieh (2008), also acknowledges the influence of personal characteristics on judgment behavior. Three levels of credibility assessment between the information seeker and information object were distinguished in interviews with undergraduate students. First, the construct level describes the users’ personal definition of credibility. This may include concepts such as truthfulness, believability, and trustworthiness. The definition of the user may deviate from the definition given by Fogg and Tseng (1999) since mental models of the construct may vary exceptionally between users due to, for instance, differences in age, education, or intelligence. The second level is labeled *heuristics* by the authors and refers to general rules-of-thumb used to estimate credibility. These heuristics include media-related and source-related heuristics. The third level concerns actual interaction with the information, which can be split into content and peripheral cues from the information itself as well as from its source.

The content and peripheral cues in the interaction level of the framework proposed by Hilligoss and Rieh (2008) is similar to the distinction between heuristic and systematic evaluation. Metzger (2007) also made this distinction in her dual-processing model of website credibility assessment. This model is strongly based on the dual-processing theory of Chaiken (1980) and predicts the type of assessment done by a user, depending on the motivation and ability to evaluate. Metzger defined heuristic evaluation as using superficial cues and systematic evaluation as constituting a thorough evaluation of a website’s credibility.

Motivation comes from the “consequentiality of receiving low-quality, unreliable, or inaccurate information online” (Metzger, 2007, p. 2087). Motivation thus can vary, as consequences of low-quality information might differ between tasks. For tasks with low importance (e.g., personal entertainment purposes), consequences of poor information could be very limited whereas tasks of higher importance (e.g., searching information for a school assignment) can have more serious consequences (e.g., a low grade). Motivation

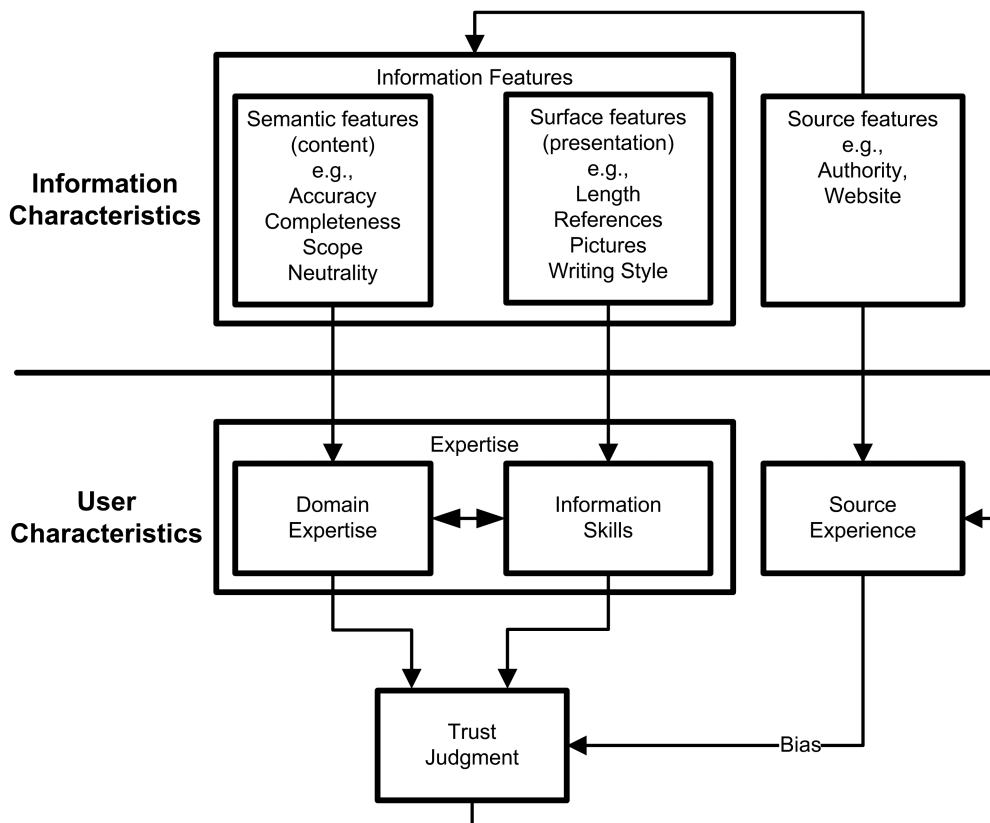


FIG. 1. The proposed 3S-model of information trust.

thus can be interpreted as the importance of credible information. When the user is not motivated, no evaluation is done at all or a heuristic evaluation is done. When the user is motivated to evaluate, however, the type of evaluation depends on the ability of the user. Ability is linked to “the users’ knowledge about how to evaluate online information” (Metzger, 2007, p. 2087). These skills can be taught to users in information skills education. If a user has the ability to evaluate, a systematic/central evaluation is done; otherwise, a heuristic/peripheral evaluation is done.

A different approach was taken by Fogg (2003). His prominence-interpretation theory predicts the impact of various noticeable elements in a piece of information on a credibility assessment. Prominence refers to the likelihood that an element is being noticed by the user. This is multiplied by interpretation, which indicates the value or meaning people assign to this element. The result is the credibility impact of the element under evaluation.

Metzger’s (2007) model mainly considers aspects of users’ motivation and ability whereas Fogg’s (2003) theory concerns the information itself without identifying aspects of the user, which may lead to different prominence or interpretation of elements. Combining the predictions of both models, one can expect that the influence of various elements in a piece of information is based on specific characteristics of a user. Metzger predicted that the type of evaluation is dependent on the ability of the user, but various levels of ability also could lead to other elements being prominent

in a piece of information. An example is the element of “references” in an article. For academic students, this is a very prominent element (Lucassen & Schraagen, 2010); however, younger school children are probably not (yet) familiar with the concept of referencing or its importance (Walraven, Brandgruwel, & Boshuizen, 2009).

This aspect of a user’s ability that Metzger (2007) described in her dual-process model is quite general. We propose to distinguish two types of expertise on the topic at hand: (generic) information skills and domain expertise. Both have the potential to influence a user’s ability to assess credibility. When a piece of information is within the area of users’ expertise, different elements are likely to be prominent as compared to information outside their area of expertise. Using elements such as accuracy, completeness, or neutrality requires knowledge of the topic at hand, which only users with a certain level of domain expertise have. However, other elements, such as the length of a piece of information or the number of references do not necessarily require domain expertise.

In this article, a new model of trust in information is proposed, as shown in Figure 1. In this model, we predict that trust judgments of a user specifically depend on the two aforementioned user characteristics: information skills and domain expertise. Based on prominence-interpretation theory (Fogg, 2003), these characteristics lead to different features in the information being used in trust judgments. Furthermore, users may alternatively choose to rely on their earlier experiences with a particular source instead of actively

assessing various features of a piece of information. In this model, we have tried to add more detail to the trust behavior of users than do current models by considering characteristics of both the user and information.

We name the proposed model the *3S-model*. The three Ss stand for *semantics*, *surface*, and *source* features of information, as well as for the three different *strategies* users may take when judging credibility of information. We discuss these three main elements of the proposed model in detail in the following sections.

Domain Expertise

Expertise has a long history in psychological research. It is well-known that experts approach problems within their domain of expertise differently than do novices. Whereas novices are known to think about problems in a concrete manner, focusing on surface characteristics, experts tend to form abstract representations, focusing on the underlying principles of a problem. For example, Chi, Feltovich, and Glaser (1981) found evidence for this difference by presenting physics problems to both experts and novices. The participants in this experiment were asked to categorize these problems into groups based on similarity of solution. Virtually no overlap was seen between the categories introduced by novices and experts. Novices tended to sort the problems according to surface features, such as the presence of a block on an inclined plane in the description of the problem. In contrast, experts generally categorized the problems into groups based on the underlying physics principles that could be applied to solve the problem.

Adelson (1984) used the same distinction between experts and novices to create a situation in which novices could actually outperform experts. Undergraduate students and teaching fellows were considered novices and experts, respectively, in the domain of computer programming. Two conditions were introduced; in the first condition, a concrete representation of a computer program was given (concerning how the program works), after which a concrete question was asked. In the second condition, both the representation and the question were abstract (concerning what the program does). The first condition should better suit novices whereas the second condition should suit experts. This hypothesis was confirmed by the measured task performance; experts were better in answering abstract questions, whereas novices answered more concrete questions correctly.

When domain experts and novices are asked to judge information credibility, similar differences to those found by Chi et al. (1981) and Adelson (1984) can be expected. When experts judge information within their area of expertise, they are able to assess the *content* on several aspects such as accuracy, neutrality, or completeness. Novices are less able to do this due to their lack of knowledge about the topic; they mainly have to rely on the assessment of *surface* characteristics.

Domain familiarity can be seen as a weaker form of domain expertise. In a think-aloud study by Lucassen and

Schraagen (2010), familiarity with the topic was varied for participants judging credibility. While no significant difference was found in the distribution of information features used, post-hoc inspection of the data showed that correctness of the information was mentioned almost solely by participants familiar with the topic. Correctness (or accuracy) of the information thus may be an important factor for trust in information, which can predominantly be judged when the user has a sufficient level of domain expertise.

Information Skills

As noted earlier, users may judge other aspects than the semantics of a text as well, such as surface features. Assessing such features does not necessarily require domain expertise; other skills are needed to identify which features are relevant to credibility. These skills can be seen as a subset of information skills. A common definition of this is “the ability to recognize when information is needed and the ability to locate, evaluate, and use effectively the needed information” (American Library Association Presidential Committee on Information Literacy, 1989). In this study, we focus on the evaluation aspect as this includes evaluation of credibility or trust. We interpret information skills as generic skills, which require no expertise in the domain of the information.

Users with varying levels of information skills approach information in different ways. Brandgruwel, Wopereis, and Vermetten (2005) investigated information problem solving by information experts (doctoral students) and information novices (psychology freshmen). The task of information problem solving was decomposed in problem definition, searching, scanning, processing, and organization of the information, all guided by a regulation process. Judging information (including credibility) is done in the information scanning and processing stages. The first stage can be seen as heuristically scanning the information whereas the latter stage involves in-depth systematic processing of the information. They found that experts put significantly more effort in the processing stage than do novices. Experts also seem to judge scanned information more often, although a difference was found only at the 10% significance level. These findings indicate differences in behavior between experts and novices in judging information, especially since their behavior was largely similar in most other stages of information problem solving.

Brandgruwel et al. (2005) further showed a difference in the amount of effort information experts and novices put into the processing of information. However, qualitative differences also can be expected. Walraven et al. (2009) for instance, showed that in group discussions by people with limited training in information skills (high-school students), many factors relevant to trust or credibility are not mentioned. Examples are objectivity and whether information comes from a primary or secondary source, but the notion of references also was mentioned only once in eight group discussions.

Lucassen and Schraagen (2010) showed that for college students, several textual features, references, and the presence of pictures were important noncontent features when judging credibility of Wikipedia articles. The differences between the importance of references for high-school students and college students can be attributed to differences in information skills. Hence, people with varying information skills can be expected to differently assess credibility of information.

We do not suggest that the strategies of employing domain expertise or information skills to form a trust judgment are mutually exclusive. Instead, we expect that for various users, strategies vary in their impact on the trust judgment. For instance, domain experts are likely to base their judgment primarily on factual accuracy whereas people with advanced information skills (e.g., information specialists, doctoral students) are likely to mostly bring to bear their information skills in their judgments when the topic at hand is out of their domain. However, it is not expected that domain experts will no longer notice surface features or that information specialists no longer notice the semantics; their domain expertise or information skills may only render certain features more prominent than others.

Furthermore, we expect that both types of user expertise interact. Consider, for example, the quality of references. Domain experts will know which journals are considered the best in their field. This knowledge can aid the information skills of the user and improve the trust judgment.

Source Experience

An alternative strategy to form a trust judgment also is introduced in the 3S-model. Instead of *actively* assessing content or surface features, the user may *passively* rely on earlier experiences with the source of the information. This behavior also was identified by Hilligoss and Rieh (2008) in the heuristics level of credibility assessment (source-related heuristics). Following this strategy, it is possible that the influence of domain expertise or information skills (and thus the corresponding features in the information) is diminished or even ruled out when a user has a lot of positive (or negative) experiences with a particular source. In this case, a user will no longer feel the need to actively judge the credibility of the information, which is similar to the prediction of Metzger (2007) that the lack of motivation leads to no assessment or a heuristic assessment.

When a trust judgment is formed following any of the three proposed strategies, this new experience is added to the preexisting experience with the source. This feedback connection also is present in the integrated model of trust in information by Kelton et al. (2008).

Heuristic Versus Systematic Processing

Using one's experience with the source of information to judge credibility can be considered highly heuristic behavior. However, semantic and surface features can be evaluated heuristically or systematically. While some of the features

listed as examples of surface features at first might seem to facilitate heuristic processing (e.g., the length of a text), surface features also can be processed systematically. An example is assessing the quality of the references: Doing this requires an effortful evaluation of each reference. The same is true for the assessment of content features: At first, this may seem to require systematic processing, but the process of comparing presented information with own knowledge can be considered recognition, which according to the RPD model (Klein, Calderwood, & Clinton-Cirocco, 1986) does not require systematic comparison of information. On the other hand, when a presented statement is just outside of the area of expertise, its validity might still be checked by bringing to bear the knowledge an expert possesses, which is typically a systematic process (resulting in the phenomenon of "fractionated expertise," described by Kahneman & Klein, 2009, p. 522).

However, we argue that assessing trust in information always will contain a certain degree of heuristics. Consider someone who systematically evaluates every single element relevant for trust in a piece of information. By doing this, the risk of using poor information is eliminated, which in itself is an important aspect of trust (Fogg & Tseng, 1999; Kelton et al., 2008). This means that trust is no longer necessary because the user has complete certainty of the credibility of the information. However, complete certainty is impossible; hence, trust assessments are always heuristic to a certain degree. Grabner-Krauter and Kaluscha (2003) also identified this in their proposition that trust and information search (systematic processing) are alternative mechanisms to absorb uncertainty. This is needed because situations are generally too complex to incorporate all relevant factors.

Hypotheses

In this study, we attempt to find empirical evidence for the validity of our proposed model, mainly focusing on the concept of domain expertise. We asked Internet users with varying expertise in one particular area (automotive engineering) to assess the credibility of Wikipedia articles on this topic. The factual accuracy of the articles was manipulated, ranging from original quality to articles containing factual errors in half of the treated concepts as well as in the topic definition. According to the proposed model, lower accuracy should affect trust judgments of users with domain expertise. This leads to the first hypothesis:

- H1:** Decreases in factual accuracy have a negative impact on trust in information of domain experts.

We hypothesize that users with little domain expertise are less able to focus on content features to assess credibility. This would mean that manipulating accuracy does not influence the trust judgments of these users, which leads to the second hypothesis:

- H2:** Decreases in factual accuracy have no impact on trust in information of novices.

These hypotheses are based on the expectation that domain experts and novices will use different cues from the article in their assessments. A substantial number of these cues can be made explicit by asking users for their rationales for their judgments (We acknowledge that some of this knowledge may be tacit and not open to verbalization.) According to the 3S-model, this leads to the final two hypotheses:

H3: Novices use surface and source features more than semantic features in trust judgments.

H4: Experts use semantic features to a larger extent than do novices in their trust judgments.

Note that the expectation that experts will use their domain expertise does not give reason to assume that they will no longer use surface features to assess credibility. This could be the case when domain experts with very limited information skills are assessing credibility, but testing such hypotheses is beyond the scope of this study.

Method

Participants

Since nearly every car brand (and model) has its own online forum with numerous members, automotive engineering was used as the domain of expertise for this experiment to easily recruit a large number of participants. Experts were mainly active at car enthusiasts' forums whereas novices were recruited mainly from other, general-purpose forums. Invitations for participation were posted on these forums, containing a link which led them to an online questionnaire. A total of 657 participants took part in the experiment (70.0% male). The average age was 27.7 years ($SD = 10.0$). We identified 317 experts and 340 novices (Definitions used for "expert" and "novice" are discussed later.) Since all participants were Dutch or Belgian (Flemish), the experiment was performed in Dutch, using articles from the Dutch Wikipedia.

Task and Procedure

The experiment was implemented in the form of an online questionnaire. When it was opened, an explanation of the experiment was provided, including an indication of its duration ("a few minutes") and the number of questions ($n = 8$). Participants were told that they would be asked for their opinion on one Wikipedia article, without specifying what aspects of the article their opinion should be about. By doing this, we made sure that the participants were not primed to specifically focus on the credibility of the article but to approach the article in a more natural manner. After reading the instructions, participants were asked to provide some general demographic information such as gender, age, and education level. On this page, they also were asked whether they worked in the automotive industry and whether they considered cars to be a hobby.

On the subsequent page, a Wikipedia article was presented. Three different articles were used in the experiment to

account for potential influences of characteristics specific for one particular article (e.g., a very lengthy article or an unusually high number of images). The topics used were "V-type engine", "Boxer-type engine", and "Twin turbo". The articles were selected to be of similar appearance (e.g., length, presence of images) and topic (car engines). Each participant viewed only one randomly selected article. It was not possible to click on the links in the article since a full-page screenshot² of the actual article was presented.

After the participants indicated that they had finished reading the article, they were asked whether they trusted it by means of a yes/no question. Next to this, a rationale for their judgment could be provided optionally. The trust question and the rationale were presented on a separate page from the Wikipedia article. To prevent multiple page views when answering the questions, it was not possible to go back to the article once the participants indicated that they had finished reading the article. The participants were made aware of this in the instructions.

To ensure that participants could fill in the questionnaire only once, IP addresses were registered, and a cookie was saved on the participants' computers. Due to the technical limitations of online questionnaires, it could not be ensured that the participants cross-checked information with other websites or visited the original page on the Dutch Wikipedia; however, none of the rationales indicated such behavior. Furthermore, we do not expect that such behavior would interfere with the goals of this study.

Independent Variables

Expertise. This variable was assessed using two questions. Participants who indicated that they worked in the automotive industry or who considered cars as a hobby were considered experts; otherwise, they were considered novices. The participants were not asked directly whether they were experts in the domain because we expected that this might lead them to read the article in a different way (e.g., especially focusing on their domain expertise). We acknowledge that this strategy of distinguishing experts from novices does not guarantee that our expert participants were absolute domain experts. However, we expect the differences in domain familiarity and expertise between our expert and novice participants to be sufficient for the purpose of this study.

Factual accuracy. This variable was manipulated by adding factual errors to the article. First, the number of concepts treated in each article was counted. Then, the facts in a predefined percentage of concepts were altered in such a manner that no inconsistencies within the article were created. Possibly due to the descriptive encyclopedic character of Wikipedia articles, there were only a few links between the concepts in one article. This means that single facts could be altered while maintaining internal consistency.

²Using Webshot, <http://www.websitescreenshots.com>

Furthermore, the facts were altered to be the opposite of the actual fact, or at least very different from it. By doing so, the presented facts were clearly incorrect. An example of an altered fact in the article on the “V-shaped engine” is the following sentence: “V-shaped engines are mostly applied in vehicles in which space is not an issue. By placing the cylinders diagonally, a V-engine takes more space than an inline or boxer engine with the same capacity.”³ Originally, the article correctly stated that these engines are applied when space is an issue because they take up less space.

The articles used were not very extensive (~600 words) and provided a brief introduction on the topic rather than an in-depth discussion. Therefore, we could assume that people with a reasonable level of domain expertise would be able to detect at least some of the errors introduced.

The manipulation was validated by showing all original and manipulated statements of each article side by side to two independent domain experts (garage owners). A substantial degree of intersubjective truth about the correctness of the statements was reached since they were able to identify the correct and incorrect statements with an accuracy of 92.3%. Only two statements were not correctly identified by the domain experts. The first statement was on the English term for a boxer engine with four cylinders (flat-four), which was incorrectly identified by both garage owners. This error likely can be attributed to a lack of proficiency in the English language. The second statement was on the angle between the cylinders and the crankshaft in a V-shaped engine, which was incorrectly identified by one of the garage owners. It is most likely that this statement was misread since this domain expert can be assumed to be highly familiar with V-engines.

The following conditions were used in the experiment:

- Original article, not manipulated
- Errors in 25% of the concepts
- Errors in 50% of the concepts
- Errors in 50% of the concepts and an error in the topic definition.

The definition of the topic is given in the first sentence of each article. It is presumably more important than the other concepts because it introduces the main concept of the article and helps to get a grasp of the subject of the article. For example, the correct definition of the “Biturbo” article is “A biturbo or twin-turbo is an internal combustion engine fitted with two turbos.”³ The manipulated definition stated that biturbo engines are diesel engines. The conditions were randomly assigned to the participants.

Dependent Variables

Trust judgment. The percentage of the participants trusting the information in the article in each condition was measured by the percentage of positive answers to the question “Do you trust the information in the article?” A dichotomous scale was

used because each participant assessed only one article. More detailed scales (e.g., a 7-point Likert scale) were considered less useful because participants could not compare articles.

Rationale for the trust judgment. The (optional) rationales for the judgments of participants were categorized into the three strategies proposed in the 3S-model: rationales based on surface features, semantic features, and source features. Rationales containing comments on multiple features were categorized according to the dominant feature in the rationale. Rationales that could not be categorized into one of these types were classified as “other”. Two experimenters both analyzed 60% of the data; Cohen’s κ was calculated for the overlapping 20%. The resulting value of 0.799 indicates a substantial agreement. A qualitative analysis of the disagreements between annotators revealed that most of them were the result of interpretation differences of the most dominant feature in rationales with multiple features.

Results

Trust Judgments

Table 1 and Figure 2 show the percentages of experts and novices trusting the information in the articles in all conditions.

The percentage of experts trusting the information decreased when the factual accuracy of the articles was manipulated, $\chi^2(3) = 7.81$, $p = 0.05$, supporting H1. For novices, no difference was found, $\chi^2(3) = 3.69$, $p = 0.30$. This supports H2.

Visual inspection of Figure 2 shows an unexpected “dip” for trust of novices in the 25% condition. Post hoc analysis showed that the number of novices trusting the information in this condition was almost significantly lower, $\chi^2(1) = 3.37$, $p = 0.066$, than that in the other conditions. Analysis of the novices’ trust judgments in this condition showed no significant differences between the three articles used, $\chi^2(2) = 1.71$, $p = 0.43$. Subsequently, a quantitative and qualitative inspection of the rationales of novices in this condition also revealed no differences compared to those in the other conditions, $\chi^2(1) = 0.15$, $p = 0.70$. Furthermore, content of the articles in this condition was examined post hoc. No unexpected irregularities in terms of, for instance, internal consistency were found.

Rationales to the Trust Judgments

A total of 520 participants (79%) gave a short rationale for their trust judgments. Table 2 gives an overview of how these rationales were divided into the categories of source features, semantic features, surface features, and other rationales.

Rationales in which the source of the information (Wikipedia) was mentioned were classified as source experience. Examples of these are “I don’t trust the information, because it’s from Wikipedia, and anyone could have put it on there, without having any knowledge on the topic” and “This

³Note that this is a translation of the original sentence; the articles used in the experiment were in Dutch.

TABLE 1. Percentages of participants trusting information in the article for varying manipulation levels. The exact numbers of experts and novices trusting the information in each condition is given in parentheses.

| | Original article | 25% errors | 50% errors | 50% errors + definition |
|---------|------------------|------------------|------------------|-------------------------|
| Experts | 82.2% (60 of 73) | 79.3% (65 of 82) | 71.3% (57 of 79) | 64.6% (53 of 82) |
| Novices | 69.8% (60 of 86) | 59.5% (47 of 79) | 69.0% (60 of 87) | 72.7% (64 of 88) |

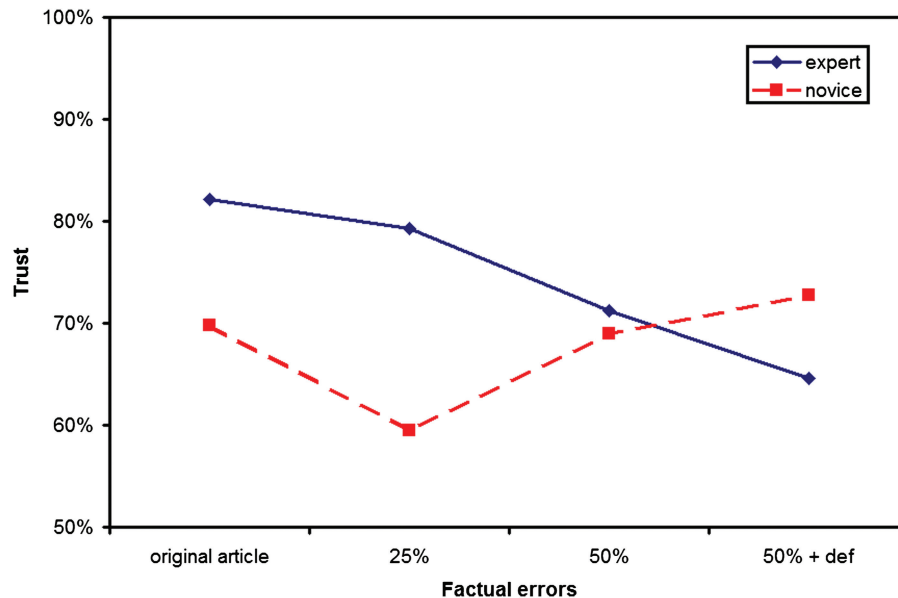


FIG. 2. Percentages of experts and novices trusting the information in each condition. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

TABLE 2. Feature categories used in trustworthiness assessments by experts and novices.

| | Experts | Novices |
|---------------------------------------|---------------------------------|----------------------------------|
| Source features (source experience) | 25.2% ($n = 60$) | 33.7% ($n = 95$) ^a |
| Semantic features (domain expertise) | 38.2% ($n = 91$) ^a | 6.7% ($n = 19$) |
| Surface features (information skills) | 32.8% ($n = 78$) | 47.9% ($n = 135$) ^a |
| Other motivations | 3.8% ($n = 9$) | 11.7% ($n = 33$) |

^aCategories with a significantly higher number of rationales by experts or novices.

is from Wikipedia, which is mostly quite accurate.” Rationales regarding factual accuracy or the preexisting knowledge of the participant were categorized as domain expertise. For instance, rationales such as “This fits with my own knowledge” and “I found some factual errors in the article” were classified as reflecting domain expertise, including rationales in which specific errors in the article were mentioned (e.g., “There never was a Boxer engine in the Alfa Romeo 156”; “This is wrong, larger turbos do not work better at lower speeds”). When surface features of the article were mentioned, the rationales were categorized as information skills.

Examples of these are “This looks well-written” and “Poor language, few references supporting the propositions.” All rationales that could not be categorized in one of these three categories were marked as “other” rationales and excluded from further analysis due to their diverse character and low percentage (8%).

The distribution of features used by novices was different from a distribution expected by chance, $\chi^2(2) = 83.66$, $p < 0.001$. Considering the low percentage of semantic features (6.7%), this supports H3, which predicted that novices would largely focus on source and surface features. The three proposed categories covered 88.3% of all rationales of novices.

The distribution of cues across the three categories (see Table 2) was different for experts and novices, $\chi^2(2) = 69.57$, $p < 0.001$. Experts used semantic features from the information more than did novices to underpin their trust judgments, $\chi^2(1) = 69.41$, $p < 0.001$, supporting H4. Moreover, novices relied more on the use of surface features, $\chi^2(1) = 19.62$, $p < 0.001$, and source features, $\chi^2(1) = 7.78$, $p = 0.005$, than did experts. The three strategies proposed by the 3S-model covered 96.2% of all rationales of experts.

Post hoc analysis showed that the source of the information was used as a rationale to both trust and not trust the information. No significant difference was found between positive

and negative use of this rationale, $\chi^2(1) = 1.09$, $p = 0.30$. Furthermore, experts and novices did not differ in their ratios of positive to negative “source” rationales, $\chi^2(1) = 0.68$, $p = 0.41$.

Discussion

The trust judgments of participants in the various conditions of this study largely support the proposed 3S-model. We found that trust of domain experts was influenced by the accuracy of the presented information. This was expected because accuracy is a key aspect of the semantic (content) features of information.

According to the proposed model, evaluating these features requires a degree of domain expertise. In contrast to domain experts, novices’ trust remained approximately the same for the various accuracy levels. The 3S-model can be used to predict that their lack of domain expertise leads them to mainly assess source and surface features, which were kept constant in this study. Hence, their trust judgments were not influenced by the manipulation of factual accuracy.

A second observation regarding trust of experts and novices is that the latter group had less trust in the information in articles of original or slightly manipulated quality. This result replicates the finding by Chesney (2006) that domain experts value Wikipedia articles as more credible than do novices. His conclusion was that Wikipedia articles are very accurate, supporting Wikipedia as a reliable information source. We hypothesize that experts have the advantage of recognizing the presented facts in the article. This gives them a very strong sense of confidence in the information since their preexisting knowledge contains the same facts. Since novices mostly lack preexisting knowledge, they do not get this sense of confirmation. When the accuracy was severely manipulated, trust of experts was similar to trust of novices, or even less. Under these conditions, the recognition of facts by experts is replaced by the recognition of errors, significantly decreasing trust.

A second explanation for the low trust of novices in comparison to experts is that “distrusting” novices are aware of their own limited abilities to judge the credibility. To avoid potential problems as a consequence of the use of poor information, novices may be highly skeptical of the information they encounter. It is trusted only when they are highly confident of the credibility. This behavior might protect them from potentially poor information, but it also may keep them from using high-quality, credible information on unfamiliar topics. This hypothesis is supported by some statements expressed by novices, such as “I don’t understand the information. A lot of terms are used which I don’t understand and which thus can be wrong”; and “I know nothing about this topic, so I am not 100% confident that this information is true.”

An important observation in the trust judgments of experts is that despite the negative influence of diminished factual accuracy on trust, the majority of the experts still trust the information. This was observed even in the condition with

the highest percentage of errors (64.6% trust of experts). This observation can be explained in several ways.

First, the experts did not exclusively use their domain expertise in the assessments. In fact, numerous rationales still referred to aspects of the source or surface features of the information. The usage of these features does not lead to variations in trust between the accuracy conditions, as source features and surface features were kept constant.

Second, the participants were intentionally not made aware beforehand that they would be asked to judge the credibility of the article. Instead, they were instructed that they would be asked their opinion, without specifying what their opinion should be about. As this was done to stimulate natural behavior on Wikipedia, it also might mean that concepts such as trust or factual accuracy were not salient to the participants. They might instead have been paying attention to other aspects of the information, such as the information load or entertainment value.

Moreover, we may expect a low motivation from the participants in our experiment since they had no personal benefit in performing well. According to Metzger (2007), this leads them to perform no evaluation at all or a heuristic evaluation. Answering the questions after viewing the article required them to do an evaluation, which thus was likely to be of a heuristic nature. Consider the large percentage of evaluations by experts based on semantic features. These evaluations might not have gone beyond swiftly skimming through the article, recognizing some of the presented facts, and inferring that all information in the article is credible based on the recognized facts. This hypothesis is supported by the fact that even in the condition with the worst factual accuracy, 50% of the treated concepts did not contain factual errors. By reading swiftly, the experts might have missed them.

A final explanation of the high number of experts trusting the information in conditions with heavily manipulated accuracy concerns their level of expertise. Our expert participants indicated that they either worked in the car industry or that they were car enthusiasts, but this does not necessarily mean that they were real experts on car engines. While the presented information was aimed at the general public rather than domain experts, this does not exclude the possibility that some of the facts were actually unknown to some of our expert participants. To find out whether our participants were able to find errors in the articles, a post hoc analysis of the rationales of experts was performed. This showed that 62% of the introduced errors were explicitly mentioned by at least one participant. However, more errors may have been found, as the participants often stated that they only found factual errors, without specifying them.

The rationales for the trust judgments of the participants in our study provide additional proof for the validity of the model. We observed that experts mainly try to bring to bear their domain knowledge with their judgments. However, this did not rule out the utilization of their source experience or information skills. In fact, numerous rationales still referred to aspects of the source or surface characteristics of the information. Following this observation, note that the use of

domain expertise, information skills, and source experience are not mutually exclusive in trust judgments; instead, a combination of these features is employed. The impact of each feature in one of the three categories depends on the characteristics of each particular user and piece of information. In this study, we observed that domain experts in automotive engineering largely used their domain expertise. Accordingly, we also predict that, for example, information specialists (e.g., librarians) will largely use their information skills and will therefore be largely influenced by various surface features (e.g., references).

Novices rarely mentioned semantic features in their rationales. This was expected because their domain expertise is at most very limited, if not completely absent. Novices mainly seem to compensate for their lack of domain expertise by assessing surface features of the information. This was reflected by a higher percentage of surface features in rationales of novices than in that of experts. Moreover, the use of source features by novices also exceeded experts' use of these features. Source experience and information skills do not require domain expertise on the topic and are thus highly accessible to novices.

As predicted by the source experience component of the 3S-model, the presented information was frequently dismissed simply because it came from Wikipedia. Experience with this particular source was clearly negative in these cases. Some participants mentioned this explicitly, whereas others referred to the underlying principles of Wikipedia (e.g., open-editing model, multiple authors). Remarkably, the source of the information also was used as a reason to trust the information, possibly because of earlier positive experiences with the website. In the case of Wikipedia, this is likely because the overall quality of Wikipedia is quite high (Giles, 2005). People who expressed this rationale did not assess content or surface features but directly gave their trust judgment based on earlier experiences with Wikipedia. These observations of the source of the information leading to a trust judgment in which the actual content of the information was not considered confirm the biasing influence of this strategy.

The limited domain expertise (of novices), which is expected in information search behavior (Lim, 2009), and limited information skills (of both novices and experts; Walraven et al., 2009) might have been the cause of the observation that users solely rely on previous experiences with the source. In most cases, this is not a problem because of the high overall information quality, but in cases when the quality of an article is disputed, users are unlikely to detect this following this strategy. Examples of such cases are vandalism or disputed neutrality (Denning et al., 2005). The high number of rationales in which the source of the information was mentioned also is a good indicator of how Wikipedia is trusted blindly by many and carefully avoided by others.

Limitations

The experiment performed in this study has brought some confirmation of the validity of the 3S-model; however, a

few limitations should be kept in mind. The participants in the experiment were recruited from online forums. While this is a great strategy to obtain a high number of participants, accountability is low. For instance, experts were distinguished from novices only on the basis of two questions prior to the experiment. We have no reason to assume misbehavior, but novices also could have posed as experts by answering these questions in a particular way. Moreover, as stated earlier, experts might have been car enthusiasts without being domain experts in car engines. This leads to the limitation that we cannot be absolutely positive that our expert participants can be considered actual domain experts. However, their level of expertise proved to be adequate for the purpose of the experiment.

Each domain of expertise will have its own specifics concerning evaluation behavior. In this experiment, we have shown that for this setting in automotive engineering, the 3S-model seems valid. However, other domains may have different specifics, potentially leading to different behavior. Examples are differences in the consequences of poor information, controversy within the domain, or education level of domain experts. The 3S-model should be investigated using other areas of expertise.

In this research, the Dutch Wikipedia has been used as a case study to provide a familiar source of information, used by numerous people. However, lots of characteristics of the information, such as the layout or the open-editing model behind it, are very specific for Wikipedia. The 3S-model should be tested on different information sources in different contexts. Both online and offline sources should be considered. User scenarios other than handling encyclopedic information also could be applied. When, for instance, health information is considered, motivation could be much higher because of the potentially high impact of the negative consequences of poor information. A second domain in which credible information is vital is the military.

The rationales for trust judgments of the participants provided valuable insights into their behavior; however, note that these could be provided optionally, and not all participants did so. This means that these results may not apply to the entire sample in this experiment.

Future Research

More empirical research into the 3S-model is necessary. Fine-grained insights into the behavior of users following the three proposed strategies and the elements of information which correspond to these strategies should be attained. This could be achieved, for instance, by conducting think-aloud experiments. Furthermore, the performed experiment focused on the manipulation of features in one of the three strategies (domain expertise). Although we have shown the employment of all three proposed strategies, future experiments also should focus on manipulating features in the other strategies (source experience and information skills).

The lack of a difference between novices' trust in credible and less credible information is an important area

to research. While this study has demonstrated this problem, more detailed (within-subject) studies should further investigate it, as this leads to novices not trusting credible information as well as novices trusting less credible information. A promising direction to address this problem is the development of support tools for information credibility. Such tools already have been researched and developed, for instance, aiming at the credibility of Wikipedia (e.g., Adler et al., 2008; Korsgaard & Jensen, 2009). The relationship between advice given by such support systems and users' own assessments should be examined. The 3S-model can provide factors to consider in such examinations. It is plausible that users who mainly use their source experience benefit more from support systems than do users who actively assess the information themselves. Differences also may be found between domain experts and novices.

Conclusion

This study has provided new insights concerning the concept of domain expertise in trust judgment behavior of Internet users. A new model of trust judgment has been proposed in which three distinct strategies are identified. Users rely on their domain expertise, information skills, or experience with the source of information to form a trust judgment. An initial validation has been performed, mainly focusing on domain expertise. More empirical studies focusing on other components of the 3S-model are necessary. Knowing these strategies, we more clearly understand how trust judgments are formed. Furthermore, we are more able to predict the information features on which trust judgments depend. The proposed 3S-model can serve as a framework for further research on trust in information and support systems.

References

- Adelson, B. (1984). When novices surpass experts: The difficulty of a task may increase with expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(3), 483–495.
- Adler, B.T., Chatterjee, K., de Alfaro, L., Faella, M., Pye, I., & Raman, V. (2008). Assigning trust to Wikipedia content (Tech. Rep. No. UCSC-SOE-08-07). School of Engineering, University of California, Santa Cruz.
- Alexander, J.E., & Tate, M.A. (1999). *Web wisdom: How to evaluate and create information quality on the Web* (1st ed.). Hillsdale, NJ: Erlbaum.
- American Library Association Presidential Committee on Information Literacy. (1989). Final report. Chicago, IL: Author.
- Brandgruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior*, 21(3), 487–508.
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39(5), 752–766.
- Chesney, T. (2006). An empirical examination of Wikipedia's credibility. *First Monday*, 11(11).
- Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121–152.
- Corritore, C., Krachera, B., & Wiedenbeck, S. (2003). On-line trust: Concepts, evolving themes, a model. *International Journal of Human-Computer Studies*, 58(6), 737–758.
- Denning, P., Horning, J., Parnas, D., & Weinstein, L. (2005). Wikipedia risks. *Communications of the ACM*, 48(12), 152.
- Flanagin, A.J., & Metzger, M.J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media Society*, 9(2), 319–342.
- Fogg, B.J. (2003). Prominence-interpretation theory: Explaining how people assess credibility online. In the *ACM Conference on Human Factors in Computing Systems (CHI '03)*, extended abstracts (pp. 722–723). New York: ACM Press.
- Fogg, B.J., & Tseng, H. (1999). The elements of computer credibility. In *Proceedings of the Special Interest Group on Computer-Human Interaction (SIGCHI) at the Conference on Human Factors in Computing Systems (CHI '99)* (pp. 80–87). New York: ACM Press.
- Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438(7070), 900–901.
- Grabner-Krauter, S., & Kaluscha, E.A. (2003). Empirical research in on-line trust: A review and critical assessment. *International Journal of Human-Computer Studies*, 58(6), 783–812.
- Hilligoss, B., & Rieh, S. (2008). Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing & Management*, 44(4), 1467–1484.
- Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, 64(6), 515–526.
- Kelton, K., Fleischmann, K.R., & Wallace, W.A. (2008). Trust in digital information. *Journal of the American Society of Information Science and Technology*, 59(3), 363–374.
- Klein, G.A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. *Proceedings of the annual meeting of the Human Factors and Ergonomics Society* (pp. 576–580). Santa Monica, CA: Human Factors and Ergonomics Society.
- Korsgaard, T.R., & Jensen, C.D. (2009). Reengineering the Wikipedia for reputation. *Electronic Notes in Theoretical Computer Science*, 244, 81–94.
- Lim, S. (2009). How and why do college students use Wikipedia? *Journal of the American Society of Information Science and Technology*, 60(11), 2189–2202.
- Lucassen, T., & Schraagen, J.M. (2010). Trust in wikipedia: How users trust information from an unknown source. In *Proceedings of the Fourth Workshop on Information Credibility (WICOW '10)* (pp. 19–26). New York: ACM Press.
- Mayer, R.C., Davis, J.H., & Schoorman, F.D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734.
- Metzger, M.J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, 58(13), 2078–2091.
- Rajagopalan, M.S., Khanna, V., Stott, M., Leiter, Y., Showalter, T.N., Dicker, A., & Lawrence, Y.R. (2010). Accuracy of cancer information on the Internet: A comparison of a Wiki with a professionally maintained database [Abstract No. 6058]. *Journal of Clinical Oncology American Society of Clinical Oncology*, 28. Retrieved from http://www.asco.org/ASCOv2/Meetings/Abstracts?&vmview=abst_detail_view&confID=74&abstractID=41625
- Walraven, A., Brandgruwel, S., & Boshuizen, H. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers & Education*, 52(1), 234–246.